

图书出版量的优化组合预测模型及其应用

王秋萍^{1,2},刘素兵¹,王晓峰¹,张军¹,曾涛¹

WANG Qiu-ping^{1,2},LIU Su-bing¹,WANG Xiao-feng¹,ZHANG Jun¹,ZENG Tao¹

1.西安理工大学 理学院,西安 710054

2.西安理工大学 工商管理学院,西安 710054

1.School of Sciences,Xi'an University of Technology,Xi'an 710054,China

2.School of Business Administration,Xi'an University of Technology,Xi'an 710054,China

E-mail:wqp566@yahoo.com.cn

WANG Qiu-ping,LIU Su-bing,WANG Xiao-feng,et al.Optimal combination forecasting model of books publication kinds and its application.Computer Engineering and Applications,2008,44(12):246-248.

Abstract: According to the historical data of the national books publication,we make use of GM(1,1) forecasting model and cubic exponential smoothing model to construct optimal combination model based on IOWGA (Induced Ordered Weighted Geometric Averaging) to forecast books publication kinds.We give test to combination forecasting model.The test result shows that this model is effective and suitable.The combination forecasting model based on IOWGA possesses high accuracy and is applicable to practice.The books publication kinds in four years,from 2006 to 2009,are predicted by established model.

Key words: book publication kinds;IOWGA operator;forecasting model;combination forecasting;Matlab

摘要:以全国图书出版量的历史数据为基础,采用灰色GM(1,1)模型和三次指数平滑模型进行组合优化,建立了基于IOWGA(诱导有序几何加权平均)算子的图书出版量组合预测模型,并对组合预测模型进行了检验。检验结果表明,组合预测模型是有效的、可靠的,具有较高的预测精度,可应用于实际预测。并利用所建预测模型对2006~2009年的图书出版量进行了预测。

关键词:图书出版量;IOWGA算子;预测模型;组合预测;Matlab

文章编号:1002-8331(2008)12-0246-03 **文献标识码:**A **中图分类号:**O211;O159

1 引言

图书是收集情报、传播知识、为社会提供信息服务的重要工具。图书的出版量一方面说明一个国家科技文化发展水平,另一方面对各级各类图书馆的文献订阅工作有重要的指导作用。因此,对图书出版情况进行定量预测,对于图书出版发行部门、各级各类图书馆、科研教育部门以及文化事业管理部门等都有极其重要的实际意义。

自J.M.Bates和C.W.J.Granger于20世纪60年代提出组合预测方法^[1]以来,组合预测方法已经成为预测领域中的一个重要研究方向,Clemen曾指出,组合预测将成为预测研究的主流之一^[2]。传统的预测图书出版量的方法有统计回归分析法^[3]、灰色GM(1,1)预测法^[4,5]、神经网络法^[6],但由于图书出版量的影响因素众多,且相互关系较为复杂和呈现非线性等特征,仅采用单一的预测模型往往使许多有用的经济信息得不到有效的利用,较难得到满意的结果。如果对单一模型进行适当的组合,则可以发挥单一模型的优点,克服单一模型的局限性,充分利用已有的信息资源,有效地集结更多有用信息,从而使预测结果具有更大的可信度。

灰色预测模型采用小样本,通过对样本数据的生成、开发,提取有用的信息,实现对系统运行规律的正确描述和有效控制。三次指数平滑法则对不同时期的数据进行非等权处理,具有时变适应性,较符合实际情况。因此,有必要对传统的预测方法进行改进,在这两种方法的基础上,建立基于IOWGA算子组合预测模型,并将其应用到图书出版量预测中。

2 单项预测模型的建立

2.1 灰色GM(1,1)预测模型

(1)数列选择

采集图书出版量最近14年(1991~2004)的数据(数据来源: <http://stats.gov.cn>,见表1前第3列)为原始数据数列。

$$X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}$$

(2)形成1-AGO数据序列

$$X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\}$$

$$x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i) \quad k=1, 2, \dots, 14$$

基金项目:陕西省自然科学基金基础研究计划(the Natural Science Foundamental Research Program of Shaanxi Province of China under Grant No.2005F02)。

作者简介:王秋萍(1964-),女,博士生,副教授,主要研究方向:预测与决策分析,系统工程;刘素兵(1980-),女,硕士生,主要研究方向:预测与决策分析;王晓峰(1966-),女,博士,副教授,主要研究方向:密码学与网络安全;张军(1980-),男,硕士生,主要研究方向:灰预测;曾涛(1982-),男,硕士生,主要研究方向:博弈模型。

收稿日期:2007-08-07 **修回日期:**2007-11-28

表1 1991~2004年全国图书出版实际值(千种)和三种预测模型预测值(千种)及预测精度

年份	序号	图书出版 量实际值	GM(1,1) 模型模拟值 x_{1t}	GM(1,1) 模型精度 p_{1t}	三次指数平滑 模型模拟值 x_{2t}	三次指数平滑 模型精度 p_{2t}	IOWGA 组合 模型模拟值 \hat{x}_{ct}	IOWGA 组合 模型精度 p_{3t}
1989	1	74.937						
1990	2	80.224						
1991	3	89.615	89.615 0	1.000 0	81.904 3	0.914 0	88.376 7	0.986 2
1992	4	92.148	86.732 1	0.941 2	95.132 2	0.967 6	93.781 8	0.982 3
1993	5	96.761	92.959 4	0.960 7	99.168 5	0.975 1	98.181 8	0.985 3
1994	6	103.836	99.633 8	0.959 5	103.336 9	0.995 2	102.755 3	0.989 6
1995	7	107.381	106.787 4	0.994 5	110.793 7	0.968 2	107.397 4	0.999 8
1996	8	112.813	114.454 6	0.985 4	114.021 6	0.989 3	114.088 5	0.988 7
1997	9	120.106	122.672 4	0.978 6	118.864 2	0.989 7	119.445 3	0.994 5
1998	10	130.613	131.480 1	0.993 4	126.636 0	0.969 6	130.719 0	0.999 2
1999	11	141.831	140.920 3	0.993 6	139.033 2	0.980 3	140.626 8	0.991 5
2000	12	143.376	151.038 2	0.946 6	152.583 2	0.935 8	151.276 1	0.944 9
2001	13	154.526	161.882 6	0.952 4	152.649 8	0.987 9	154.042 5	0.996 9
2002	14	170.962	173.505 6	0.985 1	162.975 5	0.953 3	171.833 7	0.994 9
2003	15	190.391	185.963 2	0.976 7	182.648 8	0.959 3	185.446 7	0.974 0
2004	16	208.294	199.315 2	0.956 9	206.990 3	0.993 7	205.784 3	0.988 0
		MSE	1.221 0		1.290 5		0.732 1	
		MAE	3.640 6		3.882 5		1.810 4	
		MAPE	2.887 2%		3.008 0%		1.316 0%	

(3) 建立灰色预测模型

$X^{(1)}$ 具有近似指数变化规律,白化方程为:

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = b \quad (1)$$

背景值 $z^{(1)}(k) = 0.5[x^{(1)}(k-1) + x^{(1)}(k)]$, 将式(1)离散化,微分方程变差分方程:

$$x^{(0)}(k) + az^{(1)}(k) = b \quad k=1, 2, \dots, 14 \quad (2)$$

由最小二乘估计解得(2)中参数列 $[a, b]^T = [-0.069 3, 77.546 1]^T$ 。

取 $\hat{x}^{(1)}(1) = x^{(1)}(1) = x^{(0)}(1) = 89.615$, 解式(1)得 $x^{(1)}(t) = [x^{(0)}(1) - \frac{b}{a}]e^{-at} + \frac{b}{a}$, 灰色 GM(1,1) 预测公式:

$$\hat{x}^{(1)}(k+1) = [x^{(0)}(1) - \frac{b}{a}]e^{-ak} + \frac{b}{a} = [x^{(0)}(1) + \frac{77.546 1}{0.069 3}]e^{0.069 3k} - \frac{77.546 1}{0.069 3} \quad k=1, 2, \dots \quad (3)$$

(4) 累减还原

对式(3)做 1-IGAO 累减还原, $X^{(0)}$ 的预测公式为

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) = 1 208.606 3 (1 - e^{-0.069 3}) e^{0.069 3k} \quad k=1, 2, \dots \quad (4)$$

(5) 模型检验

原始数据的关联度 $r = 0.618 3 > 0.60$, 模型关联度合格, 后验差比值 $\frac{S_2}{S_1} = 0.126 9 < 0.35$, 小误差概率 $p = 1 > 0.95$, 根据灰色相关检验标准^[9], 该模型除关联度 r 为三级指标水平外, 后验差比值和小误差概率都为一级指标水平, 可知模型为较好模型, 可以用于预测。

2.2 三次指数平滑模型

(1) 平滑值的计算

采用 1989~2004 年的数据对 1991~2004 年进行三次指数平滑模拟计算, 原始数据序列为 X_1, X_2, \dots, X_{16} , 记 $S_t^{(i)}$ 为第 t 期

第 i 次指数平滑值, α 为平滑参数, 指数平滑计算公式为:

$$S_t^{(1)} = \alpha X_t + (1-\alpha)S_{t-1}^{(1)}; S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha)S_{t-1}^{(2)}; S_t^{(3)} = \alpha S_t^{(2)} + (1-\alpha)S_{t-1}^{(3)}$$

(2) 三次指数平滑预测模型

$$\hat{X}(t+T) = a_t + b_t T + \frac{1}{2} c_t T^2 \quad (5)$$

$$a_t = 3S_t^{(1)} - 3S_t^{(2)} + S_t^{(3)}$$

$$b_t = \frac{\alpha}{2(1-\alpha)^2} [(6-5\alpha)S_t^{(1)} - 2(5-4\alpha)S_t^{(2)} + (4-3\alpha)S_t^{(3)}]$$

$$c_t = \frac{\alpha^2}{(1-\alpha)^2} (S_t^{(1)} - 2S_t^{(2)} + S_t^{(3)})$$

式(5)中 T 为滞后期, $T=1, 2, \dots$; $\hat{X}(t+T)$ 表示 $t+T$ 时刻的预测值。

由于所采用的数据有 16 个, 取初始值 $S_0^{(1)} = S_0^{(2)} = S_0^{(3)} = X_1$, α 值的选择要求选取方差最小的平滑系数, 通过计算分析, 选定 $\alpha = 0.44$ 。由于 α 的选取是按照方差最小时的平滑系数, 计算是按此要求做的, 模型可用。应用三次指数平滑模型对图书出版量的模拟结果和精度见表 1。

3 建立图书出版量的 IOWGA 组合预测模型

文[7]中给出了基于 IOWGA 算子的组合预测新方法, 现在建立基于 IOWGA 算子的图书出版量的组合预测模型。

对于图书出版量预测问题, x_{1t}, x_{2t} 分别为灰色 GM(1,1) 和三次指数平滑模型在第 t 时刻的预测值, l_1, l_2 分别为两种方法在组合预测中的加权系数。设 p_{it} 为第 i 种方法在第 t 时刻的预测精度。

$$p_{it} = \begin{cases} 1 - \left| \frac{x_t - x_{it}}{x_t} \right| & \text{当 } \left| \frac{x_t - x_{it}}{x_t} \right| < 1 \text{ 时} \\ 0 & \text{当 } \left| \frac{x_t - x_{it}}{x_t} \right| \geq 1 \text{ 时} \end{cases} \quad i=1, 2, t=1, 2, \dots, n$$

显然 $p_{it} \in [0, 1]$ 。记 $L = (l_1, l_2)^T$, 将 p_{1t}, p_{2t} 按从大到小顺序排序, 由预测精度序列 p_{1t}, p_{2t} 所产生的第 t 时刻的 IOWGA 组合预测值, 记为 \hat{x}_{ct} :

$$\hat{x}_{ct} \triangleq G_L(\langle p_{1t}, x_{1t} \rangle, \langle p_{2t}, x_{2t} \rangle) = \prod_{i=1}^2 x_{p-index(i)}^{l_i} \quad (t=1, 2, \dots, 14) \quad (6)$$

$p-index(i)$ 为第 i 个大的预测精度的下标,由表 1 给出的预测精度 p_{1t}, p_{2t} 和模拟值 x_{1t}, x_{2t} 代入式(6)计算得 IOWGA 组合预测值:

$$\begin{aligned} \hat{x}_{c1} &= G_L(\langle p_{11}, x_{11} \rangle, \langle p_{21}, x_{21} \rangle) = 89.6150 \ 81.9043^2 \\ \hat{x}_{c2} &= G_L(\langle p_{12}, x_{12} \rangle, \langle p_{22}, x_{22} \rangle) = 95.1322 \ 86.7321^2 \end{aligned} \quad (7)$$

同理可得 $\hat{x}_{c3}, \hat{x}_{c4}, \dots, \hat{x}_{c14}$ 的值。设 $e_t = \ln x^{(0)}(t) - \ln \hat{x}_{ct}$, 14 期总的组合预测对数误差平方和 S 为:

$$S = \sum_{t=1}^{14} e_t^2 = \sum_{t=1}^n (\ln x^{(0)}(t) - \ln \hat{x}_{ct})^2 = 0.0057 l_1^2 + 0.0004 l_1 l_2 + 0.0301 l_2^2$$

组合预测权重系数 l_1, l_2 通过解如下非线性规划模型得到:

$$\begin{aligned} \min S(l_1, l_2) &= 0.0057 l_1^2 + 0.0004 l_1 l_2 + 0.0301 l_2^2 \\ \text{s.t.} \quad &\begin{cases} l_1 + l_2 = 1 \\ l_1 \geq 0, l_2 \geq 0 \end{cases} \end{aligned} \quad (8)$$

通过 Matlab 优化工具箱解得 $l_1=0.8453, l_2=0.1547$, 再把 l_1, l_2 代入到 $\hat{x}_{ct}(t=1, 2, \dots, 16)$ 中计算出 IOWGA 组合值见表 1。若要进行未来 $k(k \geq 1)$ 步预测,用第 i 种预测方法最近 k 期拟合平均精度 $\frac{1}{k} \sum_{t=n-k+1}^n p_{it}$ 来反映预测区间 $[n+1, n+2, \dots]$ 上 $n+k$ 期的预测精度的大小。

4 模型分析比较及其应用

4.1 模型分析比较

为了进一步检验各模型预测效果的好坏,按照预测效果评价原则和惯例,采用了 3 项拟合误差指标: MSE(均方误差)、MAE(平均绝对误差)和 MAPE(平均绝对百分比误差)作为评判准则,对预测效果进行全方位的综合评价(表 1)。

由三种预测模型的效果比较(表 1)和三种模型的相对误差(图 1)可以看出, IOWGA 的组合预测模型的模拟精度高、预测误差小,无论从哪一种预测性能指标进行评价, IOWGA 的组

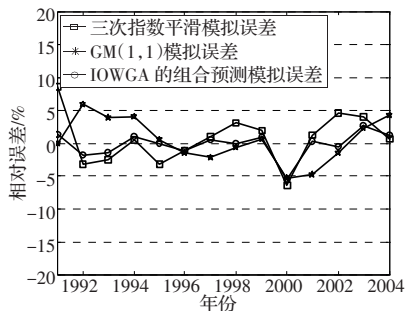


图 1 不同模型误差对比

(上接 245 页)

性。仿真结果表明,当高速公路由于某种原因发生交通拥挤时,控制系统能快速作出响应,消除交通拥挤和维持主线车流稳定。

参考文献:

- [1] Chien C C, Zhang Y P, Ioannou P A. Traffic density control for automated highway systems[J]. Automatica, 1997, 33(7): 1273-1285.
- [2] 荆便顺. 道路交通控制工程[M]. 北京: 人民交通出版社, 1995.

合预测的预测效果都要优于任意一种单一预测模型。以 2005 年的全国图书出版量对 IOWGA 组合预测结果进一步检验, 2005 年的实际值为 222.473, IOWGA 的组合预测值为 226.0175, 平均相对误差仅为 0.0159。因此, IOWGA 的组合预测模型在模拟性能和预测性能方面的改善是显而易见的。图书出版量的组合预测模型可以综合两种单一预测方法所提供的有用信息, 使考虑的因素更加全面, 因而, 较图书出版量单一预测模型而言, 它具有更好的科学性和通用性。

4.2 IOWGA 的组合预测模型的应用

利用 IOWGA 的组合预测模型预测了 2006~2009 年的图书出版量, 见表 2。

表 2 2006~2009 年全国图书出版量预测

年份	2006	2007	2008	2009
预测值/千种	226.02	249.68	294.02	320.06

5 结束语

本文将灰色 GM(1,1)模型和三次指数平滑模型, 通过优化组合, 建立了基于 IOWGA 算子的组合预测模型, 对图书出版量进行了模拟和预测。结果表明, 该模型可以作为图书出版量预测的有效工具, 并对 2006~2009 年的图书出版量进行了预测。由于图书出版受到社会、经济、文化等多方面的影响和制约, 仅以历史数据的变化规律来进行较长期的外推趋势预测, 无法使得预测结果完全精确。为此, 要结合未来的环境条件分析图书出版量的发展趋势, 从而更全面、准确地把握图书出版的发展规律, 更好地为决策提供参考依据。

参考文献:

- [1] Bates J M, Granger C W J. Combination of forecasts[J]. Operations Research Quarterly, 1969, 20(4): 451-468.
- [2] Clemen R T. Combining forecasts: a review and annotated bibliography[J]. International Journal of Forecasting, 1989, 5(4): 559-583.
- [3] 李彦萍. 对我国当代图书出版数学模型的评价与完善[J]. 图书情报工作, 1998(12): 53-54.
- [4] 程鸿飞. 灰色模型与我国期刊和图书生产规律[J]. 情报理论与实践, 1994(1): 25-27.
- [5] 王靖, 孟庆湖. 图书出版量灰色预测分析及采访对象[J]. 情报杂志, 1999, 18(1): 48-49.
- [6] 王延满, 沈思. 基于 BP 神经网络的图书发行量预测模型研究[J]. 情报杂志, 2003(6): 61-62.
- [7] 陈华友, 盛昭瀚. 一类基于 IOWGA 算子的组合预测新方法[J]. 管理工程学报, 2005, 19(4): 36-39.
- [8] 刘思峰, 党耀国, 方志耕, 等. 灰色系统理论及其应用[M]. 3 版. 北京: 科学出版社, 2004.

- [3] Khilil H K. 非线性系统[M]. 朱义胜, 译. 电子工业出版社, 2005.
- [4] 贺昱曜, 闫茂德. 非线性控制理论及应用[M]. 西安: 西安电子科技大学出版社, 2007.
- [5] Kohan R R. Decentralized nonlinear traffic control—variable structure linearizing control[C]. Proc of 43rd Conference on Decision and Control, 2004.
- [6] 梁新荣. 高速公路非线性反馈模糊逻辑匝道控制器[J]. 控制理论与应用, 2006, 2.